Electronic Textual Criticism

A Challenge to the Editor and to the Publisher

Gábor Kecskeméti

The Hungarian Academy of Sciences (HAS) was originally established in the early nineteenth century for linguistic and literary studies, including laying the foundation of research in Hungarian literary history. Studying national classics of Hungarian literary history has been the Academy's mission ever since, so HAS played and plays an essential role in the research into classic authors' life-works.[1] Scholarly text editions are made of the oeuvres of national classics or compiled from certain periods of literary history on the basis of genre. Most of these critical editions are produced by the Institute for Literary Studies of the HAS, even though faculties of literary history from prominent Hungarian universities contribute to the work. This work has been supervised by a HAS committee created for this purpose: the Textological Committee. The Committee was established in 1960 (Klaniczay) and has been operating ever since without interruption. Its functions include formulation of obligatory norms and methodological recommendations for critical text editions of Hungarian national classics; coordination and supervision of such works in different workshops; approval of plans for text editions; and accreditation of series and individual volumes as scholarly editions. In short: the Textological Committee provides quality assurance for text editions of Hungarian national classics. The efforts of the committee resulted in several hundred volumes of critical text editions in the past decades;[2] thus, the most important authors in Hungarian literary history are available for study in reliable editions of high standard that are based on carefully considered uniform principles. In the beginning, it was the publishing house of the HAS, Akadémiai Kiadó, which published all critical editions.[3] In the 1990s, four other scholarly publishing houses joined in. These critical editions provide the basis for all of the popular text editions that aim for a wider audience. These are published by several other commercial publishing houses.

So it seems that everything is alright with text editions of works of Hungarian literary history: a series of critical editions have been published with reliable text, in uniform structure, with scholarly apparatus, extensive commentaries and explanations. We would have every reason to be satisfied if the objective of the work was the same as the original goal of the Textological Committee: the production of editions in print.

However, in the twenty-first century, we cannot overlook the need for the availability of critical editions in electronic form. And the current methods and technology used in Hungary are only partially suitable for this.

Of course, a digital version can be made of any printed edition, even afterwards. In its simplest form, this is not more than a facsimile stored in a digital image format. The PDF file format became the standard in the last decade, and it is a more appropriate solution than a simple image format, because two layers can be generated in it: the facsimile appears on the user's screen, while beneath it there is a hidden character format text layer in which search operations can be performed. This text layer is usually produced automatically through optical character recognition, and usually it is not proofread even once, so its quality is not satisfying. Nevertheless, it somehow extends the user's possibilities in handling the text.

In the case of books published in the last decade – if we are lucky – there is no need to scan the facsimile or for making a character recognition. The PDF file can be produced from the publisher's file that contains the book's layout, preserving the final step of desktop editing before printing the book. This layout file has unique value: besides having a harmonious, consistent and functional format, it is usually only this file that preserves the final state of the text. Proofreading and printing approval are based on this file; final corrections and additions are usually made in this file only, so this is the only point in the technological process that records the state of the text intended for printing. From this file one can produce a single-layered PDF file that stores the text only in character form. Yet its appearance, i.e. the layout, the formats applied, the lines and pages are identical to the printed book and this text is suitable for search operations without concession.

However, the latest document accepted by the Textological Committee in 2004 calls electronic texts produced this way "digitised editions," and clearly differentiates these from "digital" or "electronic" editions (Debreczeni and Kecskeméti). This differentiation is absolutely justified and, in my opinion, we cannot aim for less in the twenty-first century than producing critical editions as truly digital editions. A digital edition is designed for the electronic medium from the very start and utilizes all the possibilities of this medium. No paper equivalent of a truly digital edition is conceivable – its "way of life," its philosophy, its principles of handling the text are all designed exclusively for the electronic environment. It offers much more complex possibilities beyond static reading of the body text and its notes. The electronic medium provides special dynamic functions of handling the text, like queries based on any criteria, functions for ordering or displaying the text in special ways. In order to open up the way for these features, the text must be structured

in an orderly fashion. The means for creating these structures are markup languages. The Textological Committee prescribes markup language structure as an obligatory requirement for electronic critical editions.

A truly digital edition cannot be created through mechanical conversion of a printed edition. The layout files produced and preserved by publishing houses are almost inadequate – even in an optimal case – for serving as bases for digital editions. The layout editor's work means a lot of added value in a paper edition, but its primary purpose goes against the clear structure of the text. Layout editors traditionally create harmonious appearance through interventions that are carried out by breaking the structural unity of the text. If there is a need, certain parts of the text receive unique direct formatting; smooth transition from line to line or from page to page is achieved by inserting 'hard' characters or breaking up paragraphs; the unity of spacing that also expresses structure is modified individually; elements that belong together functionally are separated and processed differently, etc. Even if the file before layout editing had some functional structure, this structure disappears from the layout file and the PDF file created after it. (Let us consider: the layout editor may break the functional connection between the location of reference and the footnote, and such structural connections necessarily disappear from a PDF file designed inappropriately.)

Thus, the organized technology necessary for producing digital text editions must be present from the first moment of the work in the text management practice of the specialists who work in the research process.[4] Such an undertaking cannot expect this kind of professionalism only from the publishing house, who join the work in the publishing phase. In the twenty-first century, even the research stage of a critical edition must be highly organized electronically. In a long-term endeavour that is carefully planned, employs a wider circle of contributors and specialists and aims for producing series, it is inappropriate if scholars enter their texts in general-purpose office word processors the way they can, and apply their own unique systems of notations in the process. I believe, a well-established enterprise for producing critical editions cannot exist today without a tailor-made or very highly customized specialized computer program, which means that some money needs to be invested in programming. This includes training and consultation services for the participants, which would help them make technological decisions in a unified, coherent manner. Such a text-entry application can be integrated into a database management system that records metadata for the text. Optimally, the corpus would be uploaded to a server on a monthly basis, even if parts of it are produced on

separate stand-alone computers. This means that all scholars in the project would have continuous access to the whole corpus. This would provide substantial help already in the research phase in terms of having a clear overview of the material. In the publishing phase, producing all kinds of output from the same material, from a printed book to an indexed database, would be possible. The new challenge for the publishing house is to create body text and apparatus of a specialised text edition from a complex system of databases and markup language structures of the research phase through systematically reducing them by a series of correct conversion steps.

Unfortunately, current textological practice in Hungary is very far from this. In my experience, not only individual professionals but participants of large, collective endeavours, too, use some version of Microsoft Word almost exclusively, disregarding the fact that this general-purpose word-processing application that is optimized mostly for office use does not provide all the functions indispensable for scholarly handling of text. An apparatus for critical text editions would require, for instance, the possibility of handling several different types of notes, while Word only allows two kinds; namely, footnotes and endnotes. Of course, there are several ways and degrees of utilizing the features that this program does provide and this depends on the degree of expertise. It is much better to connect automatically numbered endnotes to the references than typing the currently actual numbers and giving them upper index format manually, and then recording the corresponding notes completely independently in manually numbered paragraphs at the end of the file or in a separate file. If we miss recording the structural connection between the location in the text and the corresponding note, then we give up the possibility of achieving a near-professional state of the edition through automated conversions during a functional processing of the text. The editor of the leading periodical of literary studies in Hungary still receives manuscripts in which the line of thought is interrupted sometimes in the middle of a sentence or even in the middle of a word by some empty lines and a manually centred page number, and the sentence or the word is continued in a next paragraph. What happened was that the author who uses his or her computer as if it was a typewriter reached the bottom of the page at this point on his own display. (See Kecskeméti.)

In the conventions of recording text that result in ways of giving a clear overview of all aspects of content and form, the application of styles should have a special role. Applying paragraph styles is a rare event in Hungarian textological practice, and I have not seen an author's file in the last fifteen years in which the author used character styles for highlighting. In each and every case, standard italics, bold and other such direct formatting were used.

Individual formatting and the application of styles may be equivalent if it is only a matter of 'What will appear on the typist's display?' The moment the file enters the technological line of professional text processing there are potential dangers in individual character formatting. For example, if we achieve functional order or aesthetic unity of the text through paragraph formatting in some later stage, many applications remove individual character formatting the scope of which is more than half of the current paragraph. This way of operating the application is understandable; as it is assumed that earlier formatting, aimed at displaying the whole paragraph without expertise, is no longer needed when we produce a consistent look for the whole text through the more professional way of paragraph formatting. However, in many cases, the formatting that disappears was entered laboriously by a non-expert author and has functional meaning in the structure and interpretation of the text. Restoring such formatting needs subsequent interventions that are cumbersome and have the potential of introducing new errors. Furthermore, individual character formatting will never provide a proper basis for markup conversion. For instance, italics may have many different kinds of functions within the same text (such as highlighting titles, names, foreign words, unique expressions or quotations, or it can serve as typographical clarification, etc.). On the other hand, any number of character styles can have the same appearance, so we can create separate character styles for all the above functions, and these styles can be processed in an automated way and converted into markups. This is a fairly extensive conversion; there is a considerable added intellectual value involved. Essentially, it means transformation of the linear text into a structured database or expert system. But if we have to use Word files as points of departure, the application of styles seems like an appropriate first step on the way.

In international considerations of preparing electronic text editions and their technological realization, the usual subject of discussion is the – formerly SGML-, today XML-based – DTDs of the Text Encoding Initiative and the adaptations of these (Cover and Robinson; Sperberg-McQueen; Driscoll; Durusau; Lavagnino; Barney; in Hungarian: Zsoldos-Demjén). I am afraid that the situation is different in the field of Hungarian textology. The extent of expertise in informatics is such that colleagues who have been working on critical editions for years need to be taught how to enter correct typographic quotation marks, how to enter special characters and how to turn off autocorrecting, which makes entering critical text impossible. Hungarian textology can be proud of the decades of its book series of critical editions. But if we think we can rest content after we have published a printed edition in the belief that now the preservation of the work is ensured, then we are wrong.

We have a single layout file in the publisher's possession, which preserves the final text of the printed edition. Compared to this state of the text all other files kept by those who had worked on the text in the research process are out-of-date, since they represent the input of the process of publishing. From the layout editor's file a static representation of the text was produced in PDF. The chances are good that – because of the quick technological development – in some three years new page setting programs will not be able even to open the publisher's current file or to interpret it properly. Thus, we have not created anything that we may consider a standard electronic basis either for passing on the published text unimpaired or for using it in new ways. We have created something and we have nothing in our hands.

Just like for the whole of national cultural heritage, it is true for the texts of Hungarian literary history that creating their digital versions and publishing them on the internet must be a principal aim of national culture politics. At the same time, it is also an essential disciplinary interest. Making the texts available and searchable is basically the only tool for ensuring equal cultural and scholarly opportunities in the digital world of the twenty-first century. One of the most important challenges the Textological Committee is now facing is to care about the retrospective publication of the digitized versions of existing paper editions and, at the same time, secure that the possibilities offered by digital technology will be considered when future critical editions are prepared and published. I am convinced that in case we have reliable critical editions, we must definitely use them as the base for quality digital publishing. Their conversion needs a content exploration and semantic analysis of the text and the apparatus, the result of which is recorded according to some kind of a syntactic convention. The XML based syntactic recommendations of TEI seem to be the most widely accepted technological choices today in projects that aim at digitally processing big volumes of texts. The XML files can simultaneously ensure the dynamically generated content during the query, the support of advanced query options in the functionally formed field structure, the flexibility of the visual arrangement and formatting, and the long-term conservation and optional future conversion of the text database organized by them on a semantic basis. In my opinion, the real strength of the syntax of TEI lies in the possibility of coding the critical and interpretational apparatus together with the main text. The dynamic text generation done real time opens the way to the documentation of the text history, variants and interpretations as well, which can be easily and freely visualized functionally as needed, with the help of colours, pop-ups, typographic tools due to CSS, PHP, Javascript and further technologies. These technological solutions are absolutely suitable

for serving as a basis for a new kind of cooperation enabling the national community of professionals to join the edition of texts and the development of a new nation-wide homepage for the publication of the texts with primary importance in the literary history of Hungary. Given the fact that all the texts written in Latin are, throughout Europe, evidently included in national literary history, or in a broader sense in the textual part of national cultural heritage, these technological choices open up the possibility of cooperation even for the international community of scholars to contribute to a promising project of philological and, at the same time, digital expertise.

Notes

1. Summary of the nineteenth-century beginnings and the recent series of text editions: Szilágyi.
2. Bibliography of the critical text editions supervised by the Textological Committee can be found on the webpage of the committee (http://textologia.iti.mta.hu). This bibliography contains 461 volumes at the present time. There are only 47 items amongst them with a link to an electronic version. In other words, approximately 90 per cent of the critical text editions are available in paper editions only.
3. Akadémiai Kiadó National Company was founded after the socialization of publishing houses and presses, in 1950. The activity of the first three decades of the firm is summarized by Köpeczi. The company had come near to an economical and technological smash before the next change of the political system, by the middle of the 1980s. On this crisis as it concerned the critical text editions, see R. R.; Herman et al. The publishing house has been an incorporated company since 1996, now a co-property of the HAS and Wolters Kluwer of the Netherlands.
4. A great number of significant publications discuss the full technological process of electronic textual editing by now. For example's sake: Finneran; Sutherland; Burnard, O'Brien O'Keeffe, and Unsworth; Deegan and Sutherland; and the articles of special issue *Historical Perspectives on Digital Editing* of the periodical *Textual Cultures* (vol. 7, n. 1, 2012), especially Earhart.

Bibliography

Barney, Brett. "Digital Editing with the TEI Yesterday, Today, and Tomorrow." *Textual Cultures* 7.1 (2012) 29-41. Print.

Burnard, Lou, Katherine O'Brien O'Keeffe and John Unsworth (eds.) *Electronic Textual Editing.* New York: Modern Language Association of America. 2006. Print.

Cover, Robin C. and Peter M. W. Robinson. "Encoding Textual Criticism." *Computers and the Humanities* 29 (1995) 123-136. Print.

[Debreczeni, Attila and Gábor Kecskeméti]. "Alapelvek az irodalmi szövegek tudományos kiadásához" (Principles to scholarly editions of literary texts). *Irodalomtörténet* 85 (2004) 328-330. Print.

Deegan, Marilyn and Kathryn Sutherland (eds.) *Text Editing, Print and the Digital World.* Farnham, UK: Ashgate. 2009. Print.

Driscoll, M. J. "Levels of Transcription," in Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth (eds.) *Electronic Textual Editing.* New York: Modern Language Association of America. 2006. 254-261. Print.

Durusau, Patrick. "Why and How to Document Your Markup Choices," in Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth (eds.) *Electronic Textual Editing.* New York: Modern Language Association of America. 2006. 299-309. Print.

Earhart, Amy E. "The Digital Edition and the Digital Humanities." *Textual Cultures* 7.1 (2012) 18-28. Print.

Finneran, Richard J. (ed.) *The Literary Text in the Digital Age.* Ann Arbor: U of Michigan P. 1996. Print.

Herman József et al. "Merre tartson az akadémiai könyvkiadás?" (Where is the way to go with the book-making of the academy?). *Magyar Tudomány* 93 (1986) 467-478. Print.

Kecskeméti, Gábor. "Filológiai problémák a magyarországi irodalomtudományi kutatásban és a számítógépes kézirat-előkészítésben" (Philological problems in Hungarian literary research and in computer-prepared manuscripts), in Emil Hargittay (ed.) *Bevezetés a régi magyarországi irodalom filológiájába* (Introduction into the philology of old literature in Hungary). 3rd ed. Budapest: Universitas, 2003. 132-146. Print.Klaniczay, Tibor. "A textológiai munka problémái" (Problems of the textological work). *A Magyar Tudományos Akadémia I. Osztályának Közleményei* 19 (1962) 339-349. Print.

Köpeczi, Béla. "Az akadémiai kiadás utolsó harminc éve" (The last thirty years of the publishing activity of the academy). *A Magyar Tudományos Akadémia I. Osztályának Közleményei* 32 (1981) 45-51. Print.

Lavagnino, John. "When Not to Use TEI," in Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth (eds.) *Electronic Textual Editing.* New York: Modern Language Association of America. 2006. 334-338. Print.

R. R. "Az elnökség napirendjén: az akadémiai könyv- és folyóiratkiadás időszerű kérdései" (Some actual questions in relation to publishing books and periodicals of the academy on the agenda of the officers). *Magyar Tudomány* 93 (1986) 230-233. Print.

Sperberg-McQueen, Michael C. "Textual Criticism and the Text Encoding Initiative," in Richard J. Finneran (ed.) *The Literary Text in the Digital Age.* Ann Arbor: U of Michigan P. 1996. 37-62. Print.

Sutherland, Kathryn (ed.) *Electronic Text: Investigations in Method and History.* Oxford: Clarendon P. 1997. Print.

Szilágyi, Ferenc. "Akadémiánk 'Nemzeti Könyvtár'-terve 1850-ből s jelenlegi szövegkiadásunk" (The plan for a series with the title National Library of the Hungarian Academy of Sciences in 1850, and our recent text editions). *Magyar Tudomány* 88 (1981) 186-194. Print.

Zsoldos-Demjén Anetta. "Digitális szövegrögzítés TEI irányelvek alapján" (Digital text markup on the basis of TEI principles), in Gábor Kecskeméti and Réka Tasi (eds.) *Filológia és textológia a régi magyar irodalomban: Tudományos konferencia, Miskolc, 2011. május 25–28.* (Philology and textology in the old Hungarian literature). Miskolc: Miskolci Egyetem BTK Magyar Nyelv- és Irodalomtudományi Intézet. 2012. 491–503. Print.